734a F2F Notes 6/24/05 8:47 AM

Numerical Data
Averages = measures of central tendency:
AU - analysis unit - items of subject -
eg., 1 student in the ET program
sample = subgroup of population/ET program
(most good national are 1100 sample size)
population = every student in the ET program (70-90% doesn't require generalization)

mean - add then up/divide by the items

	Population	Sample
Mean	μ	X (with a line over it)
Standard Deviation	σ	S
Size	N	N
An observation	X	x

Sample mean x (w/ -) = sigma x/over n
Population mean mu = sigma x/over N

$\frac{n+1}{2}$ $\qquad$ $\frac{18+1}{2} = 9.5$th

(split the difference)

median - positional number (eg., prices of house) in the middle - line then up and it's right in the middle
mode - is the number that shows up the most often

range is the difference between the highest and lowest number: eg., 17 is top, 1 is lowest, therefore 17-1=16
variation - how far everyone is from the middle
dispersion - how far everyone is dispersed from the middle

percentile - how many are at your level and below (based on 100 units); (based on 10 = decile; cut into 4 pieces = quartile)

best way to show variation is with the quartile - groups of fours

17
17
17
> 16.5
16 Q3 = 3(n + 1) = 13.5th

15 4

14 ---------- middle 50% (from 3rd to 1st Quartile

14

> 12 $\frac{n+1}{2} = \frac{17+1}{2}$ = 9th - median or second quartile (Q2)

10

10

10 \ InterQuartile Range (IQR) **the range of the MIDDLE percent of the data**

10

> 10 Q1 = 1st quartile $\frac{n+1}{4} = \frac{17+1}{4}$ = 4.5th

10

9

5

1

5 - number summary

Low	Q1	Q2	Q3	high
1	10	12	16.5	450

IQR = Q3 - Q1

16.5 - 10 = 6.5

Box & wiskers chart:

```
+-----+------+-------+--------------------------------------------------------------------------------------------------+
1     |      |       |                                                                                                      450
 25% | 25% | 25% |                                          25%
      |      |       |
      +------+-------+
     10     12    16.5
```

skew is counter intuitive -

mean = # in the middle

standard deviation - how big steps are from the mean

```
                               |
                       |       |        |
               |       |       |        |       |
----+------+-------+-------+---------+-------+-------+
    6     10      14      18        22      26      30
```

example: 18 complaints - SD = 4

Probability that on any given day:

a) less than 14 complaints: 15.86%

b) between 10 - 22 complaints: 81.85%

c) more than 26 complaints: 2.27%

d) between 14 and 22 complaints: 68.26%

e) between 10 and 14 complaints: 13.59%
f) more than 22 complaints: 15.86%

using the software:
is the data normally distributed - check first two tests - if not normally distributed check "skewness"
number - negative number = negative skew
Kurtosis = measures "peaked" ness - more peak - more kurtosis -
ways to determine the normality

C.V. = coefficient of variation - relative measure simply to compare one data set to another

sampling distribution of the means....
plot all the sample means their distribution will be "normal" and their mains = population means and their standard deviation will be sigma/over/ n-squared

example: average age of pep undergrads
n (sample) = 30
x w/ line over it (mean) = 21
sigma (sd) = 2